SMS Clarification Paper

Use of Interpolation Methods for Characterization of the Distribution of Sediment Chemical Contamination, Area-Weighted Averaging, and Mass and Volume Calculations

Prepared by Maureen Goff, SAIC, for Ecology

INTRODUCTION

Washington State's Sediment Management Standards (SMS) Chapter 173-204 WAC, includes sediment source control and cleanup requirements to characterize the distribution of sediment chemical contamination and biological effects at any site of interest. The SMS rule also includes provisions in WAC 173-204-130(1) and (4) that mandate a goal of the use of latest scientific knowledge via identification, review and approval of alternate technical methods deemed appropriate by Ecology. Thiessen polygons have been a commonly used method for assigning chemical concentration values to areas between sample points at sediment sites. Ecology now considers the alternate use of interpolation methods as latest, best science to replace the use of Thiessen or other randomly assigned polygons for characterization of the distribution of sediment chemical contamination and biological effects, area-weighted averaging, and mass and volume calculations.

PROBLEM IDENTIFICATION

Thiessen polygons are created by drawing straight lines equidistantly between neighboring stations. Whole polygons are assigned the sediment chemistry concentration value of the station falling within the polygon. The area of a Thiessen polygon is solely determined by the number and configuration of station points and assigned chemistry values change abruptly at the polygon boundary.

Thiessen or other randomly assigned polygons assume neighboring sample point concentration values are independent of each other, while geostatistics prove that most environmental data are not independent. Easily accessible GIS tools are now available for more advanced and robust interpolation methods that respect and utilize the spatial relationship between neighboring environmental data points.

PROPOSED CLARIFICATION

The purpose of this clarification paper is to document the technical advantages and improved estimations of spatial interpolation methods (Inverse-Distance Weighting, Natural Neighbor, Kriging, etc) over Thiessen or other randomly assigned polygons. Using fairly simple GIS tools, interpolation methods use complex algorithms to take the influence of neighboring points into account when estimating a value at an un-sampled location. Not only do interpolation methods provide a more accurate estimate of concentrations at un-sampled locations where spatial correlation is known to exist, but they also provide gridded surfaces that allow for better delineation, mass and volume calculations, area-weighted site-averaging and cost analysis.

COMPARISON OF INVERSE-DISTANCE WEIGHTING AND THIESSEN POLYGONS IN ECOLOGY SITE CASE STUDY AND RATIONALE FOR CLARIFICATION

Technical Advantages Of Spatial Interpolation Methods

Utilization of Published Methodology for Predictions of Environmental Attributes:
Complex algorithms proven to utilize the naturally occurring spatial correlation among neighboring environmental sample points have been integrated into easy-to-use tools. GIS tools have been developed to perform and enhance methods such as Inverse Distance Weighting, Natural Neighbor, and Kriging using published methodologies for estimating values at unsampled locations. Appendix A presents the algorithm for Inverse-Distance Weighting (used in this case-study).

Kriging and IDW have added functions to improve estimations and to measure errors and uncertainty. For example, IDW uses a process called Cross Validation to iteratively compare real data points with estimated values to suggest a best cell size and best power and neighbor for a particular data set (Isaaks and Srivastava, 1989). Kriging tools often include variography as a first step in the interpolation to improve estimation by identifying the direction and extent of spatial correlation in a dataset.

Accessibility of GIS Tools:
Free tools developed by EPA FIELDS are available for geostatistical analysis (to determine the existence and extent of spatial correlation), Kriging, Inverse-Distance Weighting, Natural Neighbor, 3D visualization, and mass and volume calculations.

Interpolated grids provide more advanced analysis opportunities and functionality than Thiessen Polygons:
GIS grids created from interpolations provide added analysis and functionality, such as Area Weighted averaging, mass and volume calculations, the ability to calculate changes over time or identify trends, and visualization tools like 3D and cross-sectional viewers.

Area-weighted average calculations are simplified by the fact that all cells are a uniform size and have been assigned a concentration value. A uniform cell size means that each cell will be given the same weight and a straight mean can be used for the area-weighted average so grid statistics immediately report an area-weighted average with no further manual calculations. Thiessen Polygons require that each polygon be given a different weight in the averaging, by calculating the percent that each polygon contributes to the total area, multiplying that by the concentration, and then taking an average.

Mass and Volume calculations are performed by assigning a depth or third dimension to cells with a known area and estimated chemical concentration values.

Multiple 2-dimensional grids can be used for more complex analysis such as identifying cells where multiple conditions exist, or to find changes over time in any spatially correlated data (chemistry, bathymetry, sediment thickness) on a cell-by-cell basis. These types of analyses cannot be done with Thiessen Polygons unless all sampling events utilize all of exact same coordinates.

Comparison Of Methods

Area-weighted Averages:

Visual comparisons illustrate that thiessen polygons and grid-based interpolations such as IDW delineate areas differently. In this particular case study, a comparison of area-weighted averages and volume calculations do not indicate that one method consistently estimates higher or lower concentration values or area. There appears to be no consistent difference or relationship in the methods by these comparisons. Thiessen polygons and grid-based interpolation methods will result in different area-weighted averages, mass and volume estimates, and, ultimately, cost-projections for clean-up and disposal.

Geostatistical analysis:

In this case study, semi-variograms of datasets imply an auto-correlation of sample points for the chosen parameters from 100 to 200 feet. Semi-variograms for the datasets used in this case-study can be found in Appendix B.

Measuring Estimation Error:

A complete description of the estimation error comparison and findings can be found in the Case Study Description of the *Use of Interpolation Methods for Spatial Characterization of Sediment Chemical Contamination White Paper* prepared for the Washington State Department of Ecology (SAIC 2003). A quantitative comparison of the concentration estimation error of Thiessen Polygons and Inverse Distance Weighting was made possible by sub-setting the original datasets for Mercury, HPAH, and Acenaphthene. Data subset selection to create an initial sample event dataset and a secondary sample event dataset was based on a random sample design. A flow diagram of the process can be found in Appendix C.

The “initial sampling event dataset” was used to create both Inverse Distance Weighting interpolation and Thiessen Polygons. The “secondary sample event” subset was then used to perform estimation error analysis by comparing the estimated values of each method (based on the initial dataset) to the actual values of a secondary sampling event as a way of groundtruthing.

Mercury –	IDW error was 10% lower than Thiessen Polygons
Hpah -	IDW error was 20% lower than Thiessen Polygons
Acenaphthene -	IDW error was 94% lower than Thiessen Polygons

The Average of the Absolute Error was used to compare the accuracy of Thiessen Polygons and the IDW interpolations. An example of the Estimation Error reports can be found in Appendix D.

CONCLUSION

For the parameters used in this case study, interpolation methods proved to be a more accurate means of estimating values at unsampled sediment locations and would, therefore, give superior estimates on area, mass, volume, site-wide averages and cost. Ecology considers the best available science for characterizing sediment chemical contamination to be interpolation methods that not only respect the spatial correlation of environmental data, but also utilize the tools that provide the greatest accuracy and provide the technical advantages of working with newly developed automated tools developed for sediment characterization.

References

Burrough, Peter A. and McDonnell, Rachael A, Principles of Geographic Information Systems.1998. 98, 115-117.

EPA QA/G-5S, December 2002, *Guidance on Choosing a Sampling Design for Environmental Data Collection*, p.28.

Isaaks, E.H., and Srivastava, R.M., An Introduction to Applied Geostatistics, Oxford University Press, New York, 1989.

SAIC, 2003, *White Paper: Use of Interpolation Methods for Characterization of Sediment Chemical Contamination* prepared for Washington State Department of Ecology, Sediment Management Unit, Lacey, Washington, p. 4-9.

Watson, D.F. and Philip, G.M., *A Refinement of Inverse Distance Weighted Interpolation*, Geo-Processing, 2 .1985. 315-327.

W.Tobler, 1979, "Smooth pycnophylactic interpolation for geographical regions", *Journal of American Statistical Association*, 74, 367:519-536

Appendices

Appendix A IDW Algorithm

Appendix B Variography of Case Study datasets for Acenaphthene, Mercury, HPAH

Appendix C Data Splitting and Estimation Error Comparison Process

Appendix D Table of Estimation Error Reports (Ex: Acenaphthene)

Appendix A – IDW Algorithm

$$G(x,y)=\sum_{i=1}^{n} w_i f(x_i,y_i)$$

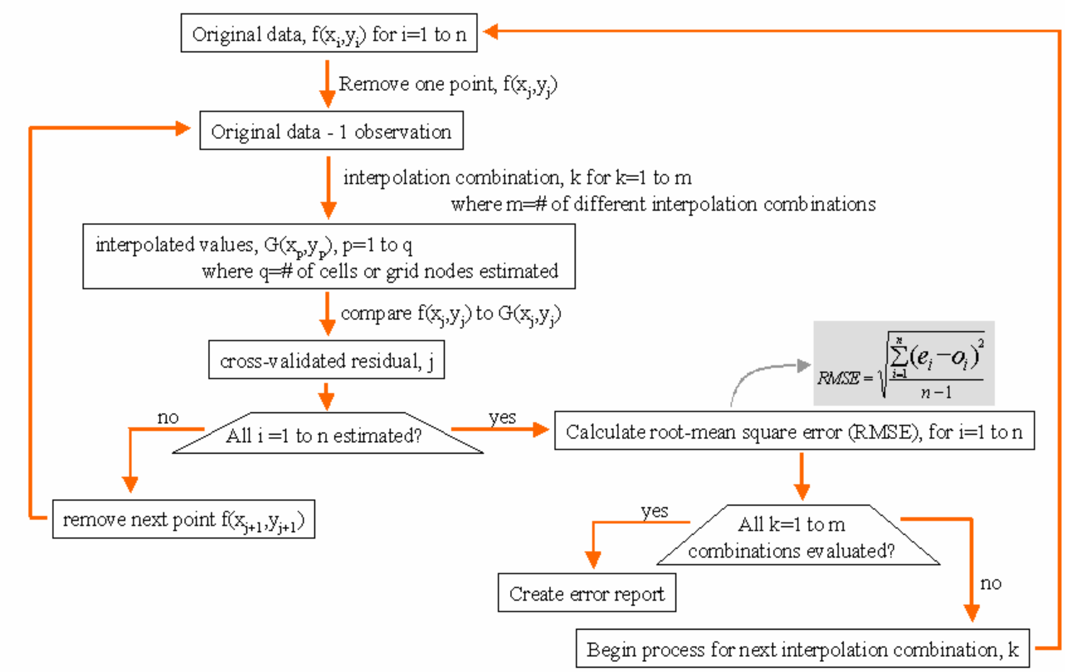
$$w_i = d_i^{-p} \Big/ \sum_j d_j^{-p}$$

where: $G(x,y)$ is the IDW estimation at (x,y);
$f(x_i,y_i)$ is the observed value at (x_i,y_i);
n is the number of nearest neighbors used for interpolation;
w_i is the weight associated with $f(x_i,y_i)$;
d_i is the distance from (x,y) to (x_i,y_i); and
p is power, a real number.

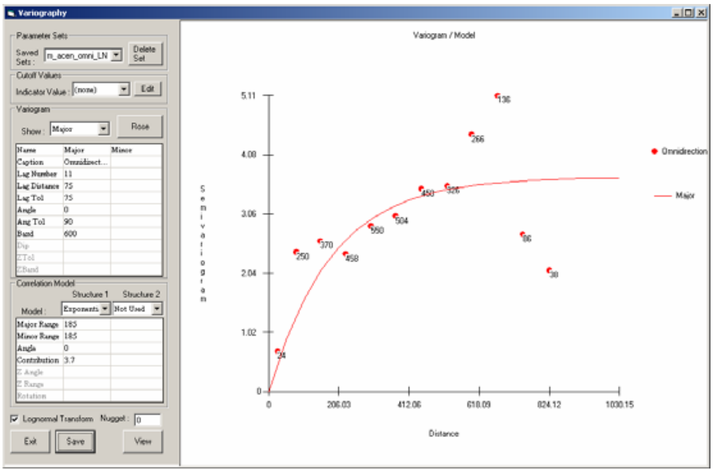
The weights are inversely related to distance and are scaled such that the sum of all the weights will add to one.

Cross Validation for IDW

Cross validation is performed for each unique interpolation permutation (e.g., neighbors =1, power=1; neighbors=2, power=1). The process and mathematical equation to find the cross-validated residual, via the root-mean square error (RMSE) equation, are provided in the following schematic.

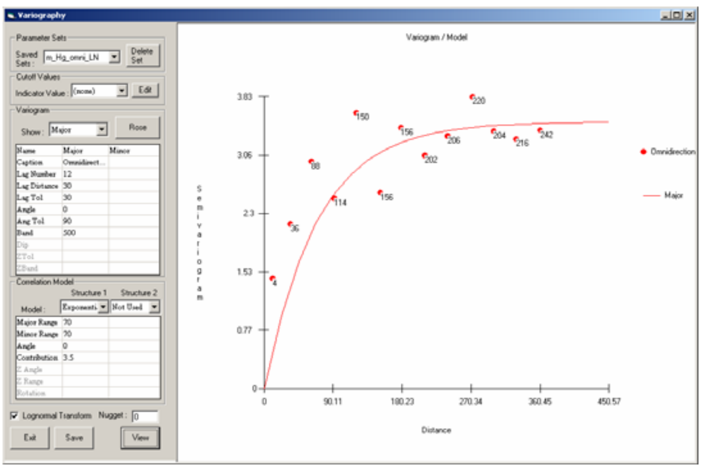


Appendix B: Variography of Case Study datasets for Acenapthene, Mercury, HPAH



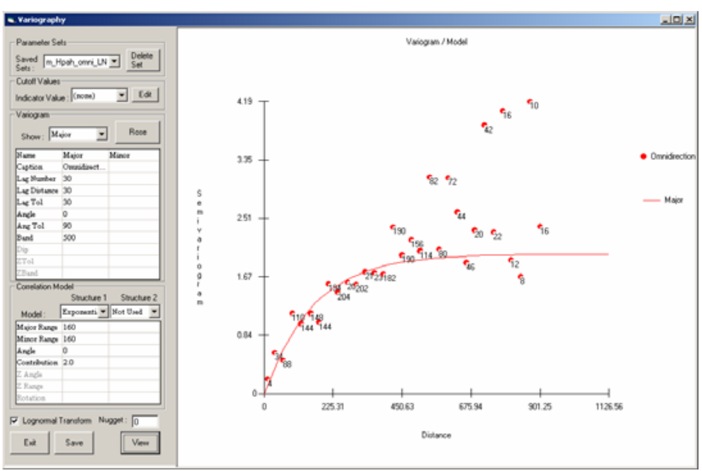
Acenaphthene -

-LN transformation
-Omnidirectional



Mercury-

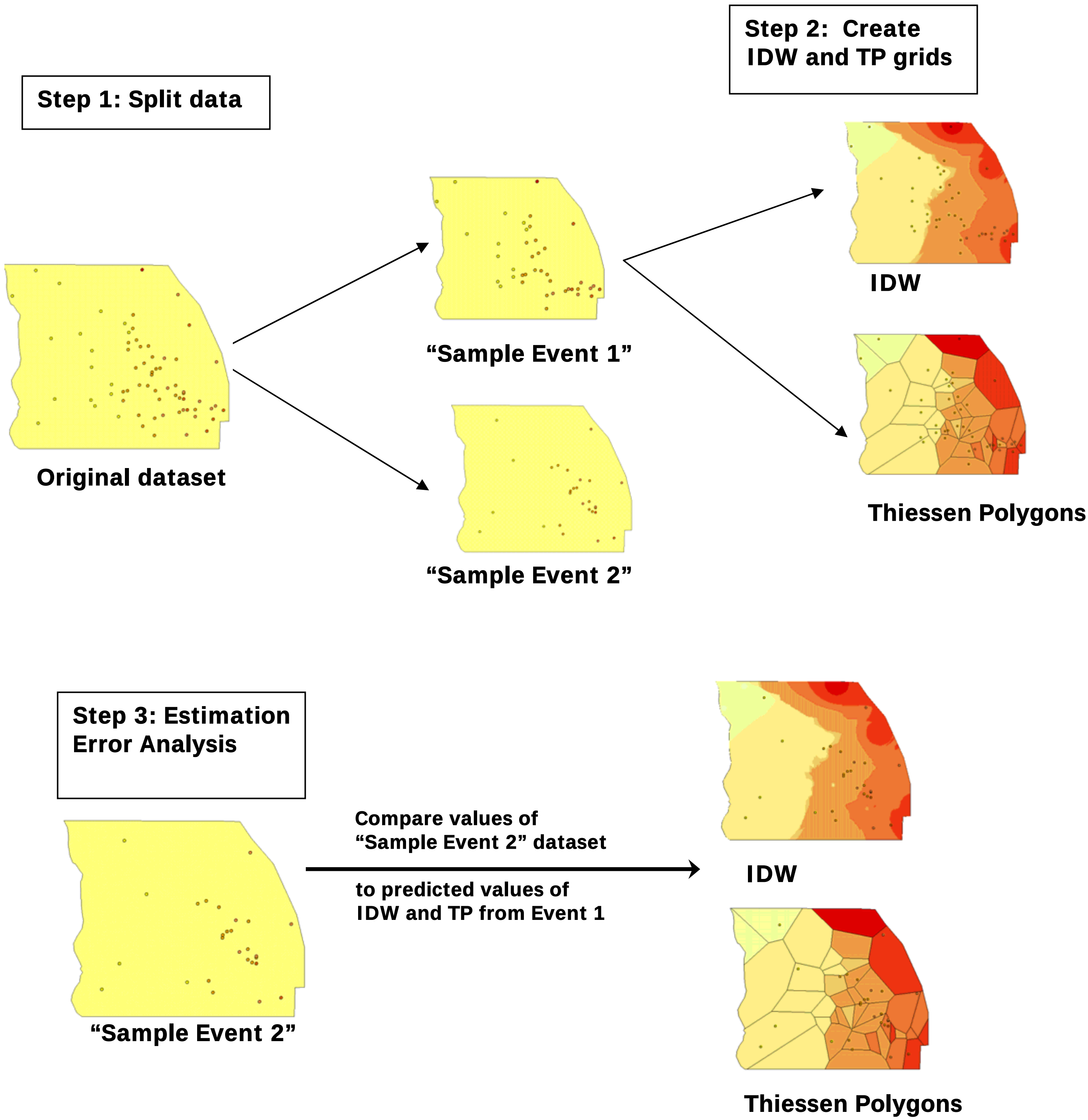
-LN transformation
-Omnidirectional



HPAH-

-LN transformation
-Omnidirectional

Appendix C- Data Splitting and Estimation Error Comparison Process



Appendix D: Table of Estimation Error Reports (Ex: Acenaphthene)

Estimated Value Error of Thiessen Polygons from Sample Event 1 comparison to Actual Values of Sample Event 2

SE2 Station ID	Actual Value	Predicted Value	Difference	Absolute Difference
WSF-1	490	310	-180	180
WSF-6	530	290	-240	240
P53VG6	1000	1300	300	300
VG-6	240	370	130	130
P53C4	190	2100	1910	1910
S11	100	1300	1200	1200
T1	50	97	47	47
S2	100	370	270	270
WSF-5	290	290	0	0
S0090	601.85	140	-461.85	461.85
T2	150	1300	1150	1150
P53VG4	34.53039	41.60959	7.0792	7.0792
SS-06	1000	700	-300	300
P53VG5	10000	310	-9690	9690
VG-5	490	310	-180	180
P53VG3	130	2100	1970	1970
S9	50	1300	1250	1250
P53VG6	59.96503	42	-17.96503	17.96503
P53VG2	560	140	-420	420
P53VG4	83	97	14	14
P53C2	1000	370	-630	630
VG-8	22	290	268	268
		Average Absolute Error		**937.9951923**

Estimated Value Error of IDW from Sample Event 1 comparison to Actual Values of Sample Event 2

SE2 Station ID	Actual Value	Predicted Value	Difference	Absolute Difference
WSF-1	490	490.25803	0.25803	0.25803
WSF-6	530	208.49883	-321.50117	321.50117
P53VG6	1000	999.64008	-0.35992	0.35992
VG-6	240	240.23918	0.23918	0.23918
P53C4	190	190.04152	0.04152	0.04152
S11	100	100.00552	0.00552	0.00552
T1	50	50.01418	0.01418	0.01418
S2	100	100.48309	0.48309	0.48309
WSF-5	290	265.46365	-24.53635	24.53635
S0090	601.85	601.84949	-0.00051	0.00051
T2	150	150.38959	0.38959	0.38959
P53VG4	34.53039	34.53066	0.00027	0.00027
SS-06	1000	999.99982	-0.00018	0.00018
P53VG5	10000	9999.66309	-0.33691	0.33691
VG-5	490	490.1265	0.1265	0.1265
P53VG3	130	130.11105	0.11105	0.11105
S9	50	590.60797	540.60797	540.60797
P53VG6	59.96503	77.53046	17.56543	17.56543
P53VG2	560	559.99994	-0.00006	0.00006
P53VG4	83	83.01068	0.01068	0.01068
P53C2	1000	999.94153	-0.05847	0.05847
VG-8	22	362.69498	340.69498	340.69498
		Average Absolute Error		**56.69734364**

-93.95547609 **Percent Difference in Average Absolute Error**